

Uso de metadatos Dublin Core en la descripción y recuperación de artículos de revistas digitales

Assumpció Estivill	Ernest Abadal	Jorge Franganillo	Jesús Gascón	J. M. Rodríguez Gairín
Universitat de Barcelona	Universitat de Barcelona	Universitat de Barcelona	Universitat de Barcelona	Universitat de Barcelona
Tel: +34 93 4035773	Tel: +34 93 4035887	Tel: +34 93 4035784	Tel: +34 93 4035788	Tel: +34 93 4035777
Fax: +34 93 1035772	Fax: +34 93 4035772	Fax: +34 93 4035772	Fax: +34 93 4035772	Fax: +34 93 4035772
estivill@ub.edu	abadal@ub.edu	franganillo@ub.edu	gascon@ub.edu	rodriguez.gairin@ub.edu

Abstract:

El trabajo resume una experiencia de aplicación del conjunto de metadatos Dublin Core a artículos de revistas digitales. Se apuntan algunos de los problemas que se han presentado a lo largo del proceso.

Keywords:

Metadata, Dublin Core, Journal articles.

1. Introduction

La revista *BiD: textos universitaris de Biblioteconomia i Documentació*¹ fue probablemente una de las primeras revistas españolas en formato digital que incluyó el conjunto de metadatos Dublin Core (DC) en la cabecera de la revista, en la de cada número y en la de todos los artículos. Con toda seguridad, fue la primera revista digital española del ámbito de la biblioteconomía y la documentación que siguió esta práctica de manera exhaustiva.² Poco a poco se ha ido incrementando el uso de este conjunto de metadatos por parte de algunas revistas españolas. Por ejemplo, *Cuadernos de documentación multimedia* incluye metadatos DC en los artículos de su último número, y desde el portal Temària³ se trabaja en la asignación de estos metadatos a los artículos de revistas digitales españolas especializadas en biblioteconomía y documentación. Asimismo, el proyecto e-revist@s del portal Tecnociencia de la Fundación Española para la Ciencia y la Tecnología (FECYT), que acoge revistas científicas electrónicas de España y de la América Latina, establece como uno de los criterios de selección recomendados —aunque no sea un requisito *sine qua non*— la presencia de metaetiquetas DC.⁴ Sin embargo, el uso de

estos metadatos no es todavía una práctica habitual de la mayoría de las revistas que se publican en el web. Títulos de larga tradición digital, con una orientación temática hacia las tecnologías de la información y la organización de la información digital y que han sido una plataforma para difundir el estándar DC —como *D-Lib magazine* o *Ariadne*—, ellas mismas no parecen haber apostado por su adopción.⁵ Este trabajo analiza la experiencia del uso del conjunto de metadatos DC en la revista *BiD* y en el portal Temària de revistas digitales españolas de biblioteconomía y documentación, y reflexiona sobre las ventajas y los inconvenientes de esta práctica.

2. La asignación de metadatos DC a la revista *BiD* y al portal Temària

Inicialmente, y ya desde el número 4, de junio de 2000, *BiD* empezó a incluir metadatos DC en la cabecera de cada una de las unidades de la revista. El objetivo principal era experimentar con el uso de estas metaetiquetas. Se trataba, sobretudo, de hacer efectivo un compromiso con los estándares y no tanto una cuestión de efectividad para mejorar la recuperación de los artículos de la revista: la mayoría de los buscadores no reconocían las etiquetas DC,⁶ y tampoco las reconocía el buscador interno de la revista (AtomZ)⁷. Así es que, para hacer efectiva la recuperación de los artículos, se tuvieron que incorporar también metaetiquetas específicas sobre las cuales pudiera actuar aquel buscador. Posteriormente, se desarrolló una base de datos organizada según los elementos del DC. Como resultado de esta aplicación, los metadatos dejaron de formar parte de la cabecera de las unidades de la revista. Aun así, posteriormente se ha recuperado esta práctica y actualmente los

metadatos se integran en la base de datos y en la cabecera de cada uno de los 255 artículos de *BiD* que se han publicado hasta el momento.

A partir de esta experiencia, el 2004 se empezó a trabajar en la creación de Temaria, un portal que tiene el objetivo de facilitar la consulta de artículos de revistas españolas de biblioteconomía y documentación mediante el conjunto de elementos DC. Actualmente este portal indiza exhaustivamente las revistas siguientes: *Anales de documentación*, *BiD: textos universitarios de biblioteconomía y documentación*, *Cuadernos de documentación audiovisual*, *Cuadernos de documentación multimedia* e *Hipertext.net*. Está previsto incorporar otros títulos en el futuro inmediato. Actualmente se pueden consultar unos 650 artículos.

Temaria dispone de una *Guía de procedimientos para la aplicación de metadatos* que concreta la aplicación del estándar DC y da directrices detalladas sobre la elaboración del contenido de los elementos. La guía tiene como objetivo asegurar la asignación correcta y uniforme de metadatos.⁸

Trece de los quince elementos del DC se han definido como obligatorios y repetibles, y se aplican a la descripción del artículo tal como se detalla en la tabla 1. En el proyecto Temaria el uso del elemento “Coverage” es opcional, y se ha descartado el elemento “Source”.

Elemento	Contenido
title	Título del artículo
creator	Autor o autores del artículo
subject	Materia del artículo
description	Resumen del artículo
publisher	Editor de la revista
contributor	Colaboradores (editores literarios y traductores)
date	Fecha de publicación del número de la revista que incluye el artículo
type	Tipo de recurso (siempre es “text”)
format	Formato del artículo (por ejemplo, html, pdf)
identifier	Identificador del artículo (en general su url, pero también podría ser otro identificador, como un DOI o un SICI) Citación bibliográfica: título de la revista y designación numérica y cronológica del número que contiene el artículo.

Elemento	Contenido
language	Lengua del artículo
relation	Relación del artículo con la revista que lo contiene y con otros recursos
rights	Información sobre el copyright, la propiedad intelectual y la protección jurídica del artículo.

Tabla 1. Definición del contenido de los elementos del DC en el proyecto Temaria

Temaria ha optado por calificar los elementos del DC. Siempre que es posible, se utilizan los calificadores definidos en el documento *DCMI metadata terms*.⁹ Pero también se usan calificadores locales más adecuados al entorno del portal. La tabla 2 resume los calificadores utilizados en el proyecto.

Elementos DC	Refinamientos de los elementos	Esquemas de codificación
creator publisher contributor	personalName* corporateName* address*	—
subject	—	TBD ¹⁰ CDU2000
description	abstract tableOfContents	—
date	issued	W3C-DTF
type	—	DCMI Type Vocabulary
identifier	—	URI (URL, DOI, SICI...)
	bibliographicCitation	—
language	—	ISO 639-2
relation	isVersionOf hasVersion isPartOf isFormatOf hasFormat	URI (URL, DOI, SICI, ISBN, ISSN, etc.)
coverage	spatial	ISO 3166 MARC 21 (043)

Elementos DC	Refinamientos de los elementos	Esquemas de codificación
	temporal	W3C-DTF

Tabla 2. Sumario de los calificadores del proyecto Temaria

* Refinamientos locales

Repasando los quince elementos del DC, probablemente para la mayoría de las aplicaciones la distinción entre “Creator” y “Contributor” es poco operativa. Si lo que se pretendía era acercar el estándar a la práctica bibliotecaria, no se ha conseguido, ya que no hay un paralelismo entre estos elementos y las etiquetas 1XX y 7XX del formato MARC. De cara a la recuperación no tiene mucho sentido hacer esta distinción, máxime cuando la misma definición del elemento contributor es poco precisa y puede llevar a ambigüedades y confusión.

A lo largo de todo el desarrollo del conjunto de metadatos DC, el elemento “Source” ha sido interpretado de maneras muy distintas. Las diferentes versiones de la guía actual *Using Dublin Core* tampoco han sabido concretar y ejemplificar el uso de este elemento que, seguramente, debería formar parte del elemento “Relation”.

Finalmente, el elemento “Coverage” tampoco tiene mucho sentido en un conjunto que se pretende muy general y de uso amplio. De hecho, este concepto está incluido en el elemento “Subject”, ya que se trata de la cobertura espacial y/o temporal del contenido del recurso.

En lo que respecta a los calificadores, en el proyecto se ha intentado usar los imprescindibles. Así por ejemplo, en los elementos “Date” y “Relation” sólo se han definido aquellos calificadores que son de utilidad en el proyecto. El uso de los calificadores refleja las decisiones que se han ido tomando con relación al contenido de los elementos, como la adopción relativamente reciente del *Tesoro de biblioteconomía y documentación*. En algún caso se han conservado calificadores que se desarrollaron inicialmente pero que, de momento, no han demostrado su utilidad —como los refinamientos “personalName” y “corporateName” que complementan los elementos “Creator”, “Publisher” y

“Contributor”. En cambio, al ser Temaria un portal que reúne revistas del ámbito académico es posible que en futuro se desarrolle algún mecanismo para anotar la afiliación de los autores y enlazarla con el autor correspondiente.

Uno de los problemas principales de la aplicación del conjunto de metadatos DC a la descripción y recuperación de artículos de revista, es la lentitud con que se ha desarrollado una fórmula para incluir la cita bibliográfica de la revista que incluye el artículo —imprescindible, por otro lado, para aplicar el DC a este tipo de recursos. Si bien en el seno de la Dublin Core Metadata Initiative funciona, desde 1998, el Citation Working Group, que tiene como objetivo desarrollar un mecanismo para incluir información sobre la citación bibliográfica de los recursos, lo cierto es que, después de diversas propuestas que finalmente han sido sucesivamente descartadas o rechazadas, no ha habido una recomendación definitiva hasta junio de 2005.¹¹

La asignación de metadatos a la revista *BiD* y a las revistas del portal Temaria también ha ido reflejando en lo relativo a la cita bibliográfica de la revista, las soluciones que derivaban de la bibliografía sobre el tema y también las apuntadas por el Citation Working Group a lo largo de los años en que ha estado activo. Las distintas soluciones de anotar aquel elemento necesario en la descripción y recuperación de artículos de revista han pasado por los estadios siguientes:

- Uso del elemento “Source”.
- Desarrollo de un campo local que no aparecía en los metadatos públicos del artículo.
- Uso del calificador “bibliographicCitation” en el elemento “Identifier” desarrollado según las directrices del documento *DCMICite: a bibliographic citation Dublin Core structured value (DCSV) encoding scheme*¹² que finalmente fue rechazado por el Usage Board de la DCMI.
- Uso del calificador “bibliographicCitation” en el elemento “Identifier” según la última propuesta del Citation Working Group.¹³ De momento no se utiliza la opción de anotar la citación bibliográfica según el estándar Open URL, sino que se anota de forma textual.

El elemento “Relation”, con el calificador “isPartOf”, se ha usado siempre para relacionar

el artículo de la revista con el título de la revista a partir de su ISSN.

2. Conclusiones

El conjunto de elementos DC tuvo un primer momento de aceptación en la elaboración de algunos portales temáticos, la mayoría de los cuales partían de proyectos relacionados con bibliotecas. Sin embargo, a diez años de su creación y después de que haya sido aprobado como norma ISO (ISO 15836:2003), no parece que su adopción esté generalizada. Un estudio reciente sobre el uso del DC en 100 proveedores de datos registrados en la Open Archives Initiative muestra que sólo 82 de los proveedores tienen metadatos disponibles para su análisis, que sólo cinco de los quince elementos del DC —“Creator”, “Identifier”, “Title”, “Date” y “Type”— están presentes en el 71% de los casos, y que 44 de los 82 proveedores solo usan los elementos “Creator” y “Identifier” en la mitad de sus productos.¹⁴ A todo ello se podría añadir que el uso de los cinco elementos anteriores son tan obvios que difícilmente pueden atribuirse a la existencia del estándar DC.

Algunos autores son muy críticos con este conjunto de metadatos y con el desarrollo lento que ha experimentado.¹⁵ Ciertamente, si se compara la aceptación que ha tenido el DC a diez años de su creación y el nivel de desarrollo alcanzado con la aceptación y desarrollo de un formato tan complejo como el MARC en el mismo período de tiempo, la conclusión es tan obvia que no hace falta comentarla.

Por lo que respecta a la aplicación de los metadatos DC a la revista BiD y al portal Temaria, la conclusión es que su uso ha facilitado un marco normalizado para el proyecto. También ha permitido que los metadatos de la revista pudieran incluirse automáticamente en la base de datos e-revist@s del portal Tecnociencia después de adaptarlas al Protocol for metadata harvesting de la Open Archives Initiative. El aspecto más negativo es que todo el proceso haya estado presidido por numerosos titubeos derivados de falta de concreciones en algunos elementos y en la ausencia de directrices firmes para la inclusión de la citación bibliográfica de los artículos.

References

1. *BiD: textos universitaris de biblioteconomia i documentació*. <<http://www.ub.es/bid/>>.
2. Carme Cantos Puig. Les metadades a les revistes electròniques espanyoles de biblioteconomia i documentació. *BiD: textos universitaris de Biblioteconomia i Documentació*. desembre, núm. 11. <http://www2.ub.es/bid/consulta_articulos.php?fichero=11cantos.htm> [Consulta: 09-04-2005].
3. *Temaria: revistes digitals de biblioteconomia i documentació*. Facultat de Biblioteconomia i Documentació, Universitat de Barcelona, Grup de Recerca Organització i Recuperació de Continguts Digitals, 2005. <<http://bidoc.ub.es/temaria/>>. [Consulta: 10-04-2005].
4. *E-revist@s: selecció i criteris de qualitat de les revistes que se incorporen al projecte*. FECYT: CSIC. <<http://www.tecnociencia.es/e-revistas/proyecto/proyecto4.jsp>>. [Consulta: 10-04-2005].
5. Una búsqueda por “Dublin core” en la revista *Ariadne* (<http://www.ariadne.ac.uk/>) recupera 385 noticias; realizada en *D-lib magazine* (<http://www.dlib.org/dlib.html>), recupera 316. [Consulta: 06-04-2005].
6. Robin Henshaw, Edward Valauskas, “Metadata as a catalyst: experiments with metadata and search engines in the Internet journal, First monday”. *Libri*, vol. 51, no. 2 (2001), p. 86–101.
7. Atomz. <www.Atomz.com>.
8. *Guia de procediments per a l'assignació de metadades Dublin Core en el marc del projecte Temària, revistes digitals de biblioteconomia i documentació*. <<http://bidoc.ub.es/temaria/lilibrestil.php>>. [Consulta: 10-04-2005]
9. *DCMI metadata terms*. DCMI, 10-01-2005. <<http://dublincore.org/documents/dcmi-terms/>>. [Consulta: 10-04-2005].
10. *Tesaurus de biblioteconomia i documentació* (Madrid: CINDOC, 2002).
11. *Guidelines for encoding bibliographic citation information in Dublin Core metadata*. DCMI, 13-06-05. <<http://epub.mimas.ac.uk/DC/dc-citation-guidelines/>>. [Consulta: 10-07-2005].
12. Ann Apps, *DCMICite: a bibliographic citation*

-
- Dublin Core structured value (DCSV) encoding scheme.* 21 Aug. 2003.
<<http://epub.mimas.ac.uk/DC/citdesv.html>>.
[Consulta: 10-04-2005].
13. *Guidelines for encoding bibliographic citation information in Dublin Core metadata.* DCMI, 29-03-05. <
<http://epub.mimas.ac.uk/DC/dc-citation-guidelines/>>. [Consulta: 10-04-2005].
 14. Jewel Ward (2004), "Unqualified Dublin Core usage in OAI-PMD data providers". *OCLC systems & services*, vol. 20, no. 1, p. 40-47.
 15. Véase, por ejemplo, Jeffrey Beall (2004), "Dublin Core: an obituary". *Library hi-tech news*, no.8, p. 40-41.